VER
Vascular & Endovascular Review

Towards Precision Cardiology through Computational Modeling: Comparative Analysis of ML-Based Predictive Models

Sanjeev Gour[1], Hemant Pal[2], Rajdeep Singh Solanki[3], Karunanidhi Pandagre[4], Rajendra Randa[5], Shiv Shakti Shrivastava[6]

[1]Asst. Professor, Dept. Of Computer Science, Medicaps University Indore, India. **
[2]Asst. Professor, Dept. Of Computer Science, Medicaps University Indore, India.
[3]Asst. Professor, Dept. Of Computer Science, Medicaps University Indore, India.
[4]Associate Professor, Dept. of MCA, Bansal Institute of Science & Tech. Bhopal, India.
[5] Research Scholar, Dept. Of Computer Science, Medicaps University Indore, India. *
[6] Professor, Parul Institute of Technology, Parul University, Vadodra, India.

Corresponding authors:
*Rajendra Randa - rajenranda001@gmail.com
**Sanjeev Gour- sunj129@gmail.com (ORCID: 0000-0002-2164-1590)

ABSTRACT

Heart issues are one of the major and serious health issues that are affecting the mortality rate at worldwide level. Having heart-related problems can also increase the chances of having other minor and major health issues. Heart issues do not depend on age and gender factors, and they can occur at any age. In the age of advanced technology, such as ML, DL, AI, NLP, LLM and so on, are widely used in various domains such as banking, marketing, education, healthcare and so on. We used supervised learning to predict any diseases in their early stage because we have values with specific parameters and features in medical data. In this study, we used various machine learning methodologies such as Linear and Logistic Regression, DT, Random Forest, KNN, Support Vector Classification and Gradient Boosting. This study focused on a comparative analysis of heart failure chances using supervised machine learning. In future studies, this work can improve the quality of results more robustly by decreasing false positive values using advanced methods, such as deep learning, NLP, and BERT methods.

KEYWORDS: heart disease, predictive healthcare, clinical text data, machine learning.

How to Cite: Sanjeev Gour, Hemant Pal, Rajdeep Singh Solanki, Karunanidhi Pandagre, Rajendra Randa, Shiv Shakti Shrivastava., (2025) Towards Precision Cardiology through Computational Modeling: Comparative Analysis of ML-Based Predictive Models, Vascular and Endovascular Review, Vol.8, No.11s, 27-31.

INTRODUCTION

Machine Learning is one of the core domains of Artificial intelligence and Advanced technology, and it is used in healthcare to achieve significant outcomes. Machine learning can analyze complex medical records to predict diseases, find patterns from unseen data, identify diseases by symptoms, structure the records and also can predict diseases in early stages, analyse data, and recognise patterns from medical records. Using advanced technology like ML and dl, we can easily find solutions to complex problems from structured and unstructured medical data. ML play an important role in the healthcare industry, and ML can also help to find out how we can solve any predictive problem in a better way using algorithms. ML and related technologies are used in data analysis and comparison between unstructured values to identify diseases and patterns of symptoms. This study. In this paper, we used a supervised model on clinical text data related to heart failure.

LITERATURE REVIEW

There are several studies have assessed the application of ML, DL and AI models in healthcare, particularly for diagnosing cardiovascular diseases and improving predictive accuracy. For instance, a study (1) evaluated ML/DL models such as vanilla CNN, FCNN, and RestNet for right ventricle (RV) segmentation in cardiac magnetic resonance imaging (MRI). These models help scan MRI images, detect RV function, and identify anomalies, enabling early consultations with physicians. However, the research suggested that combining features from both RV and left ventricle (LV) could enhance the accuracy of heart disease predictions. Another study (2) proposed AI-based prediction models using logistic regression and random forest algorithms for heart disease detection. While the research highlighted the utility of Python libraries and advanced ML algorithms, it lacked an in-depth discussion on the complex random forest algorithm and emphasised the need for more transparent and explainable AI models in this context. The use of AI-powered health chatbots has also been explored in a study (3) that developed a general architecture combining ML and natural language processing (NLP) techniques like natural language generation and understanding. The chatbot examined health status using client data, core NLP engines, and APIs. However, the study identified challenges in finding suitable ML/DL models for handling complex patient databases while maintaining data privacy and medical accuracy. A systematic literature review (4) investigated the explainability of supervised ML models in healthcare, emphasising techniques like SHAP and GRAD-CAM. Furthermore, AI-based risk prediction tools are now being trialled in healthcare systems. According to The Guardian (2024), the NHS in England is testing an AI tool to assess the risk of fatal heart disease, demonstrating the growing acceptance of ML models in real-world settings. Despite the promising applications of ML in heart failure prediction, challenges such as data quality, model generalizability, and ethical considerations remain. Roth et al. (2017) underscored the global burden of cardiovascular diseases, highlighting the need for diverse datasets to improve model robustness. Additionally, the interpretability of complex deep learning models remains a challenge, necessitating the integration of explainable AI techniques. Furthermore, AI-based risk prediction tools are now being trialled in healthcare systems. According to The Guardian (2024), the NHS in England is testing an AI tool to assess the risk of fatal heart disease, demonstrating the growing acceptance of ML models in real-world settings. Despite the promising applications of ML in heart failure prediction, challenges such as data quality, model generalizability, and ethical considerations remain. Roth et al. (2017) underscored the global burden of cardiovascular diseases, highlighting the need for diverse clinical records to improve model robustness. The study noted a significant increase in ML model accuracy but indicated that ML methods often outperform AI for large-scale healthcare applications. In a 2023 study (5), hybrid ML and NLP models, including logistic regression, AdaBoosting, and linear discriminant analysis, were applied to predict acute coronary syndrome (ACS). This research integrated structured and unstructured data to build a diagnostic framework but highlighted the difficulty of predicting unstructured data accurately using existing ML, DL, and NLP models. A novel study [6] introduced a stacking ensemble learner (SEL) alongside models like KNN, logistic regression, decision trees, support vector machines, and random forests to predict emergency readmissions for heart disease patients This research emphasized training and testing methods for both labelled and unlabeled datasets, suggesting that incorporating transfer learning, reducing training data, and leveraging diverse datasets could further enhance model reliability.

Additionally, the interpretability of complex deep learning models remains a challenge, necessitating the integration of explainable AI techniques. Future research should focus on developing hybrid models that combine traditional techniques with emerging advanced technology for enhanced accuracy and transparency. Moreover, interdisciplinary collaborations between clinicians and data scientists are essential for ensuring the practical implementation of ML-based heart prediction tools. The application of ML in heart failure prediction has shown significant promise in enhancing early diagnosis, risk stratification, and prognosis. While challenges remain, ongoing advancements in deep learning, NLP, and feature selection techniques continue to improve the accuracy, compatibility and reliability of predictive models. As AI integration in healthcare progresses, machine learning will likely play a crucial role in transforming heart failure management and improving patient outcomes.

METHODOLOGY

This study uses a comparative analysis approach to predict whether they have heart diseases or not based on clinical text data. And for it, we are done using the following parts:

Dataset: The dataset sample in figure 1, is taken for heart prediction to compare machine-learning models for predicting heart issues in its early stages. In the taken dataset, we have features e.g. Age, gender, impulse, pressure height, pressure low, glucose, kcm, troponin and class (result) holding 1320 patient records the class feature is used to predict the outcomes using various algorithms with having ‘positive’ and ’negative’. And that defines binary values as a result.

```
   age  gender  impluse  pressurehight  pressurelow  glucose    kcm
troponin  \
0   64       1       66            160         83.0    160.0   1.80
0.012
1   21       1       94             98         46.0    296.0   6.75
1.060
2   55       1       64            160         77.0    270.0   1.99
0.003
3   64       1       70            120         55.0    270.0  13.87
0.122
4   55       1       64            112         65.0    300.0   1.08
0.003

      class
0  negative
1  positive
2  negative
3  positive
4  negative
```

Figure 1: Description of clinical dataset

In our dataset, we have 2 classes, positive and negative to represent the state of heart failure in medical terms. They depend on various features such as the age of patients, gender, impulse of the heart, blood pressure, KCM, troponin and level of glucose. These are major factors to recognize heart failure.

Data Analysis and Preprocessing: we classified class features with positive and negative conditions of the heart. 80% of the overall records we used for training the model, and the rest of the data was used in testing and validating the models to improve the performance. In the used dataset, we have some null values but we have enough size of data. So, we can remove records with null values. We performed the following preprocessing steps. Such as Records with null values were removed to maintain dataset integrity. To normalise numerical features with standardization, we applied under feature scaling process. The Synthetic Minority Over-Sampling Technique is used to address class imbalances in order to balance the classes.

Machine Learning: Machine learning provides a comprehensive framework for effectively utilising various algorithms to solve complex analytical and classification problems. In our study, we aimed to compare well-known machine learning algorithms to evaluate their compatibility with medical datasets, specifically focusing on heart disease prediction. We employed Linear Regression, Logistic Regression, Decision Tree, RF, Support Vector Machine, GB classifier and KNN to analyse medical text datasets. The classification task was binary, categorizing patients into positive and negative classes—where "positive" indicates the presence of a health condition, and "negative" signifies the absence of any health issues.

Prediction: We analyze the model's performance using classification reports from different models to predict heart diseases using binary classification from clinical text data. As a result, we define the model and then train and test the model and find the result in the form of accuracy and classification reports.

Experimental- Setup:
To complete this study, we have to import different frameworks for different tasks such as importing data, analyzing and processing the data and importing, defining and predicting the unseen data. There are many frameworks for it, some of them are:

- data importing and analysis: for data importing and analysis, the data is used with 'NumPy' and 'pandas'. And for data visualisation, we used various libraries such as 'matplotlib', 'seaborn' and 'Plotypy'.
- data preprocessing: to preprocess the data we used 'pandas' to structure the data, removing the false values, and filling the values after scaling the features.
- machine learning: to use machine learning, we used the 'sci-kit learn' framework with various features such as 'ensemble', 'tree', 'matrix' and so on.

These features and frameworks provide a sufficient environment to complete the study and evaluate the models perfectly to get significant outcomes. By including some extra features, we can apply some advanced models based on ML, DL and also ensemble techniques.

RESULTS:

Each model's performance is evaluated using evolution parameters such as accuracy, precision, recall, and F1-score. The classification report provides additional insights into the true positives, true negatives, false positives, and false negatives. There are various models with their accuracy given in table form and also in graphical representation using bar chart. We can easily compare the compatibility of each model with others based on their accuracy.

Model	accuracy	precision	recall	F1-score
LR	0.681818	-	-	-
LG	0.765152	0.76/0.77	0.71/0.81	0.73/0.79
DT	0.982424	1.00/0.99	0.98/1.00	0.99/0.99
RF	0.982424	0.99/0.99	0.99/0.99	0.99/0.99
SVM	0.727273	0.73/0.73	0.63/0.81	0.68/0.76
KNN	0.640152	0.61/0.66	0.57/0.70	0.59/0.68
GB	0.982424	0.99/0.99	0.99/0.99	0.99/0.99

Table 1: Machine Learning predictions for heart failure

In this table, there is an accuracy with their evolution model features to provide a complete analysis on the compatibilities in the heart failure clinical dataset. In our study, we found that the linear model is not a better fit for the heart failure dataset because the linear model is better for continuous data, and our data prediction based on classification, but other models, such as logistic, SVM, KNN and so on, are better fits for a classification problem. RF, DT and gradient boosting provide better compatibilities in the early stage of disease prediction.

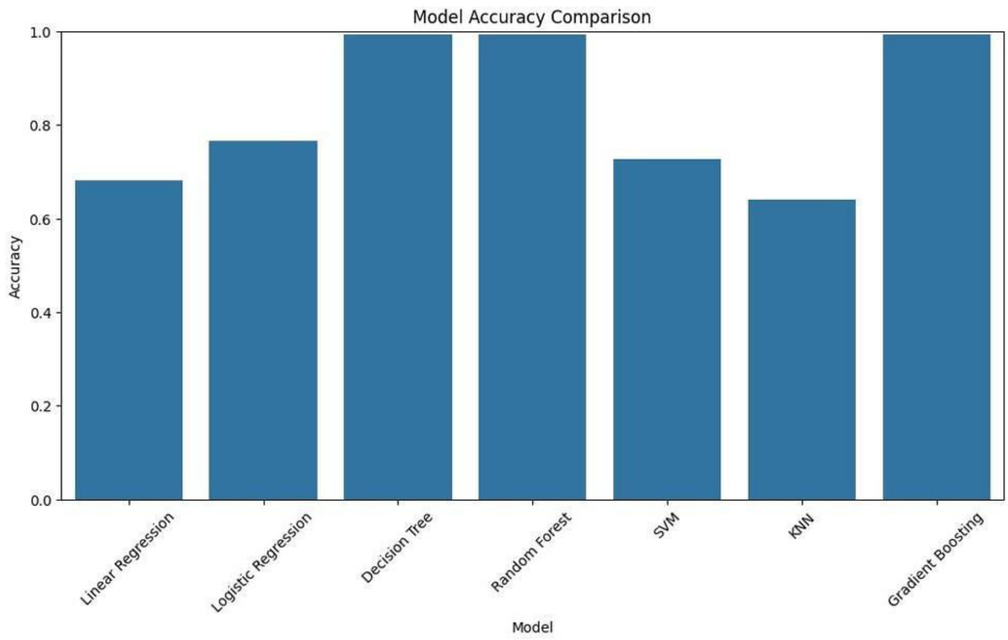

Figure 2: Graphical representation of comparison of learning models

Figure 2. is a graphical representation of the performance of this study during the experiment. Heart failure is one of the common health issues but sometimes it can lead to death in the absence of correct treatment at the right time. We used machine learning models to predict heart failure robustly which can improve the quality of medical treatment.

DISCUSSION:

This study evaluated and compared machine learning for heart failure using patient health data. The results demonstrate significant variations in performance among different models, highlighting the importance of selecting an appropriate algorithm for medical diagnosis.

The results show that DT, RF, and Gradient Boost classifiers achieved the highest accuracy (98.24%). These models excel in learning complex patterns within the dataset, making them highly effective for this classification task. The high recall and precision values indicate that these models can correctly identify both positive and negative cases with minimal errors. Logistic Regression, with an accuracy of 76.52%, performed well but was outperformed by ensemble-based methods. While it provides a good baseline model for classification problems, its performance may be limited by its assumption of linear decision boundaries. Support Vector Machine (SVM) achieved an accuracy of 72.73%. Despite its ability to handle complex classification tasks, its performance in this study was lower than that of ensemble models, which suggests that SVM may not be the best choice for this particular dataset. K-Nearest Neighbors (KNN) had the lowest accuracy at 64.02%, likely due to its sensitivity to data distribution and the course of dimensionality. It is possible that optimizing the value of k and feature scaling methods could improve KNN's performance. Linear Regression was used as a benchmark but, as a regression model, it is not ideal for classification tasks. After converting its output into binary predictions, it achieved 68.18% accuracy, which is significantly lower than other classification models.

CONCLUSION:

Applications of machine learning and artificial intelligence are increasing as increasing the complexity of data in the field. By comparing the accuracy scores and other metrics, you can determine which machine learning algorithm performs best for predicting heart attacks in your dataset. Logistic Regression, Random Forest, and Gradient Boosting Classifiers are often good starting points for classification problems. As future work, this study remains to fine-tune hyperparameters for better performance by exploring deep learning techniques using TensorFlow/Keras and Evaluating the models on additional metrics like computational efficiency.

REFERENCES:

1. Bhan, A., Mangipudi, P., & Goyal, A. (2023). An assessment of machine learning algorithms in diagnosing cardiovascular disease from right ventricle segmentation of cardiac magnetic resonance images. Healthcare Analytics, 3(1).
2. Chang, V., Bhavani, R., Xu, A., & Hossain, M. (Year). An artificial intelligence model for heart disease detection using machine learning algorithms. Healthcare Analytics, 2(1).

3. Achtaich, K., Fagroud, F., Zahra, & Naceur, A. (Year). AI-Powered Health Chatbots: Toward a general architecture. Procedia Computer Science, 191, 355-360.
4. Allgaier, J., Mulansky, L., Draelos, R., & Pryss, R. (2023). How does the model make predictions? A systematic literature review on the explainability power of machine learning in healthcare. Artificial Intelligence in Medicine, (413).
5. Emakhu, J., Etu, E., Monplaisir, L., Aguwa, C., Arslanturk, S., Masoud, S., Tenebe, I. T., Nassereddine, H., Hamam, M., & Miller, J. (2023). A hybrid machine learning and natural language processing model for early detection of acute coronary syndrome. Healthcare Analytics, 4(1).
6. Ghasemieh, A., Lloyed, A., Bahrami, P., Vajar, P., & Kashef, R. (Year). Title of the Paper. Decision Analytics Journal, 7(1).
7. TMora, T., Roche, D., & Rodríguez-Sánchez, B. (Year). Predicting the onset of diabetes-related complications after a diabetes diagnosis with machine learning algorithms. Diabetes Research and Clinical Practice, (204).
8. Allgaier, J., Mulansky, L., Draelos, R., & Pryss, R. (2023). How does the model make predictions? A systematic literature review on the explainability power of machine learning in healthcare. Artificial Intelligence in Medicine, (413).
9. Emakhu, J., Etu, E., Monplaisir, L., Aguwa, C., Arslanturk, S., Masoud, S., Tenebe, I. T., Nassereddine, H., Hamam, M., & Miller, J. (2023). A hybrid machine learning and natural language processing model for early detection of acute coronary syndrome. Healthcare Analytics, 4(1).
10. Siddiqui, S., Ahmad, A., & Fatima, N. (2021). IoT-based disease prediction using machine learning. Computers and Electrical Engineering, (108).
11. Rao, S., Li, Y., Ramakrishnan, R., Hassaine, A., Canoy, D., Cleland, J., Lukasiewicz, T., Salimi-Khorshidi, G., & Rahimi, K. (2021). An explainable Transformer-based deep learning model for the prediction of incident heart failure. arXiv preprint arXiv:2101.11359.
12. Haque, E., Paul, M., & Tohidi, F. (2024). Predicting heart failure with attention learning techniques utilizing cardiovascular data. arXiv preprint arXiv:2407.08289.
13. Wang, J. (2021). Heart failure prediction with machine learning: A comparative study. Journal of Physics: Conference Series, 2031(1), 012068.
14. Mansur Huang, N. S., Ibrahim, Z., & Mat Diah, N. (2021). Machine learning techniques for early heart failure prediction. Malaysian Journal of Computing, 6(1), 10-20.
15. Zhou, L., Guo, Z., & Wang, B. (2021). Risk prediction in patients with heart failure with preserved ejection fraction using gene expression data and machine learning. Frontiers in Genetics, 12, 652315.
16. Zhao, H., Li, P., & Zhong, G. (2022). Machine learning models in heart failure with mildly reduced ejection fraction patients. Frontiers in Cardiovascular Medicine, 9, 1042139.
17. Zhao, X., Sui, Y., & Ruan, X. (2022). A deep learning model for early risk prediction of heart failure with preserved ejection fraction by DNA methylation profiles combined with clinical features. Clinical Epigenetics, 14, 11.
18. Shin, S., Austin, P. C., & Ross, H. J. (2021). Machine learning vs. conventional statistical models for predicting heart failure readmission and mortality. ESC Heart Failure, 8(1), 106–115.
19. Sharma, V., Kulkarni, V., & McAlister, F. (2022). Predicting 30-day readmissions in patients with heart failure using administrative data: A machine learning approach. Journal of Cardiac Failure, 28(5), 710–722.
20. Sun, Q., Jiang, S., & Wang, X. (2022). A prediction model for major adverse cardiovascular events in patients with heart failure based on high-throughput echocardiographic data. Frontiers in Cardiovascular Medicine, 9, 1022658.
21. Tripoliti, E. E., Papadopoulos, T. G., & Karanasiou, G. S. (2017). Heart failure: Diagnosis, severity estimation and prediction of adverse events through machine learning techniques. Computational and Structural Biotechnology Journal, 15, 26–47.
22. Yancy, C. W., Jessup, M., & Bozkurt, B. (2017). 2017 ACC/AHA/HFSA focused update of the 2013 ACCF/AHA guideline for the management of heart failure. Circulation, 136(6), e137–e161.
23. Metra, M., & Lucioli, P. (2020). Corrigendum to 'Prevalence of heart failure and left ventricular dysfunction in China: The China Hypertension Survey, 2012-2015'. European Journal of Heart Failure, 22(4), 759.
24. Roth, G. A., Johnson, C., & Abajobir, A. (2017). Global, regional, and national burden of cardiovascular diseases for 10 causes, 1990 to 2015. Journal of the American College of Cardiology, 70(1), 1–25.
25. Bupa to offer first genetic test for disease prediction in UK. (2025, February 24). The Times.
26. NHS in England to trial AI tool to predict risk of fatal heart disease. (2024, October 23). The Guardian.